HH-FRBC: Halving Hierarchical Fuzzy Rule-Based Classifier

Azam Amouzadi
Electrical and Computer Eng. Dept, Isfahan University of Technology
Isfahan, Iran
a.amouzadi@ec.iut.ac.ir

Abdolreza Mirzaei
Electrical and Computer Eng. Dept, Isfahan University of Technology
Isfahan,Iran
mirzaei@cc.iut.ac.ir

Abstract— **The main goal of this paper is to improve the accuracy of Mamdani fuzzy rule-based classification systems. Although these kinds of systems tend to perform successfully in the case of interpretability, they suffer from rigid pattern space partitioning. Consequently, a new hierarchical fuzzy rule-based classifier based on binary-tree decomposition is proposed in this paper with the aim of creating a more flexible pattern space partitioning. The Decomposition process is controlled by fuzzy entropy of each partition. In addition final rule sets obtained by the proposed method are pruned to overcome the over fitting problem. Actually, the performance of the proposed method is compared with some fuzzy and non-fuzzy classification methods on a set of bench mark classification tasks. The experimental results show principally the good behavior of the proposed algorithm.**

Keywords—component; Keywords: Mamdani fuzzy rule-based classification systems; hierarchical fuzzy rules; fuzzy entropy

I. INTRODUCTION

Fuzzy rule-based systems (FRBSs) are valuable computational intelligence based tools, which are often used for system modeling [1]. The main interest in using FRBSs arises from the fact that these systems allow us to deal with the imprecise, noisy, or even incomplete information, which usually exists in most of the real world problems. In addition, fuzzy rules are able to describe nonlinear input/output relationship[2, 3].

Working with FRBSslead to the existence of two different types of system modeling including linguistic modeling represented by Mamadani FRBSs and fuzzy modeling represented by Takagi-Sugeno-Kang (TSK) FRBSs. According to the importance of the interpretability or the accuracy, one of the Mamdani or Takagi-Sugeno methods is used. The salience of Takagi-Sugeno system appears clearly when the accuracy increases, whereas Mamdani system should be used when more attention is drawn to interpretability. Since higher accuracy is preferred in this paper, we use Mamdani-type fuzzy rule-based systems with the purpose of improving its accuracy for the classification purpose. The generic structure of a Mamdani FRBSs comprises two main components: The knowledge base (KB) which stores the available knowledge about the problem in the form of fuzzy "IF- THEN" rules and a fuzzy reasoning method (FRM), which classifies a new sample with the information given by the KB[4].The KB includes two different parts, i.e., a rule base (RB) and a data base (DB). One of the significant problems related to Linguistic Mamdani FRBSs is the inflexibility of the DB, which imposes hard restrictions to the fuzzy rule structure. This drawback consider as a loss in accuracy when modeling some complex systems[5]. In fact, if too coarse fuzzy partitioning is used, because of generating overgeneral fuzzy rules, the performance of classification system may decrease. In the case that we use too fine fuzzy partitions, many of the created fuzzy subspaces will have no or few training samples. The rules created for such subspaces are too specific and will cause the system to over fit the training samples. Accordingly, extending the usual structure of DB to make it more flexible is a possible way to develop the performance of linguistic modeling without losing interpretability to a higher degree. To achieve this goal, Cordon et al.[6] presented a refinement approach to the simple linguistic modeling. As a matter of fact, the authors of this method introduced the concept of "layer" to extend the knowledge Base (KB) structure of linguistic FRBSs. In this extension, the KB which is called hierarchical knowledge base (HKB), is composed of hierarchical database (HDB) and hierarchical rule base (HRB). HKB formed by a set of layers, where each one defines linguistic partitions with different granularity levels for DB which causes to generate HDB and linguistic rules whose Linguistic variables take values from HDB. It should be emphasized that this method is introduced as a regression method. However this approach is usually employed for classification of imbalanced dataset and general classification in [7] and [8] respectively. The fundamental drawback of this approach seems that the number of obtained fuzzy if-then rules becomes enormous as the dimension of the problem space increases because when a rule in a typical layer is selected to be expanded with finer granularity, all of its variables become finer. The objective of this paper is to extract certain rules via an appropriate approach which includes simultaneous support for fine and coarse subspace and it also can prevent creating the large number of rules.

. To serve this purpose, a new hierarchical structure is recommended through which rules are produced in different levels with different sizes based on data distribution. Based on that, in the new hierarchical structure in each level some rules will be chosen to expand based on their entropy. After that, each selected rule will be spitted along only one of its variables based on the data distribution. At last, a pruning process will be done to avoid over fitting. To prove accuracy of the proposed method, its performance is compared to a couple of non fuzzy classifiers including ILGA [9] and LDWPSO [10]. Also the main purpose of this paper is to analyze the effect of hierarchical structure, by comparing its performance with the one obtained by the simple Chi algorithm in standard classification in [11] and . HFRBCS[7]. Furthermore, the performance of the new classifier is statistically compared using some non-parametric tests.

This paper is set up as follows. Next section introduces the components of Mamdani FRBCS and a brief description of fuzzy entropy which is used in this paper. Section 3 gives a detailed description of proposed hierarchical classifier. In section 4 experimental analysis will be included. Section 5 summarizes the paper.

II. PRELIMINARIES

In this section, the components of Mamdani FRBCS, including KB[1] and FRM[2] will be described. As it was mentioned before, KB comprises two parts, DB and RB which will be explained separately. Also a brief description of fuzzy entropy will be discussed in this section.

A. The Data base of Mamdani fuzzy rule-based classification system

A *data base* (DB), contains the linguistic term sets which are used in the linguistic rules and the membership functions defines the linguistic term sets. Each linguistic variable in the problem will be associated with a fuzzy partitioning. Fig.1 shows an example of a fuzzy partitioning which comprised of four triangular-shaped fuzzy membership functions along with their labels.

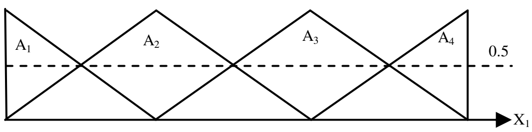

Fig. 1. example of a fuzzy partitioning

B. The Rule base of Mamdani fuzzy rule-based classification system

A *rule base* (RB) contains a set of linguistic rules that are joined together to make a good decision. Each rule is formed by the linguistic term sets which are defined in DB. Rules can be generated in different types[10], but in this paper, rules of the following type are used.

[1] Knowledge base
[2] Fuzzy reasoning method

Let us assume that we have m training (i.e., labeled) patterns $x_p = (x_{p1}, \ldots, x_{pn})$, $p = 1, 2, \ldots, m$ from M classes where x_{pi} is the ith attribute value ($i = 1, 2, \ldots, n$) of the pth training pattern. So a typical rule For FRBCS, can be defined as:

$$\begin{aligned}&\text{Rule Rj}: \text{if } x_1 \text{ is } A_{j1} \text{ and} \ldots \text{and } x_n \text{ is } A_{jn} \text{ then}\\&\text{Class} = C_k \text{ with } CF = CF_j\end{aligned} \quad , (1)$$

where Rj is the label of the jth rule, A_{ji} is an antecedent fuzzy set, C_k is the class label, and CF_j is the certainty factor of jth rule [11]. According to[12], if the antecedent fuzzy sets of the rule R_j are known, the consequent class and its CF can be determined by the following steps:

1. Computing the compatibility grade of each training pattern x_p**,** with the fuzzy if-then rule Rj, by the product operator as

$$\mu_{Rj}(x_p) = \mu_{A_{j1}}(x_{p1}) \times \ldots \times \mu_{A_{jn}}(x_{pn}) \; , \quad (2)$$

where $\mu_{Aji}(.)$ is the membership function of the antecedent fuzzy set A_{ji} and "×" is used as T-norm.

2. Computing the sum of compatibility grades for each class as follows:

$$\beta_{\text{Class}\, h}(\text{Rj}) = \sum_{x_p \in \text{Class}\, h} \mu_{Rj}(x_p) \qquad h = 1, \ldots, M \; , (3)$$

3. Finding the consequent class $C_j$that has the maximum value of $\beta_{\text{Class h}}(R_j)$ among the M classes:

$$\beta_{\text{Class}\, C_j}(\text{Rj}) = \max\{\beta_{\text{Class}\,1}(\text{Rj}), \beta_{\text{Class}\,2}(\text{Rj}), \ldots, \beta_{\text{Class}\,M}(\text{Rj})\} \; , (4)$$

4. Specifying the certainty factor of the rule Rj by (5), after determining consequent class *Cj* by (4)

$$CF_j = \beta_{\text{Class}\, C_j\,(\text{Rj})} - \bar{\beta} \Big/ \sum_{h=1}^{M} \beta_{\text{Class}\, h\,(\text{Rj})} \; , (5)$$

Where

$$\bar{\beta} = \sum_{h \neq C_j} \beta_{\text{Class}\, h}(\text{Rj}) \Big/ C - 1 \quad (6)$$

C. Fuzzy reasoning method

As it was mentioned before, an FRM is an inference procedure which classifies a new sample with the information given by the KB. There are different methods forreasoning, which are presented in [10]. In this paper the weighted vote method is used as an inference method in which, to classify an input pattern *xp* by *S*, which is a set of rules, first, the compatibility grade of x_p with the antecedent part of each rule is calculated through Eq.2. Then the strength of the vote given by each rule is defined as the product of compatibility grade and certainty factor. After that, the total strength of the vote for each class is calculated by:

$$\sigma_{\text{Class } h}(xp) = \left\{ \sum_{\substack{R_j \in s, \\ C(R_j)=\text{Class } h}} \mu_j(xp) CF_j \right\}, (7)$$

Finally, the test pattern x_p will be classified as the class having maximum total strength.

D. Fuzzy Entropy

The fuzzy entropy which is used in this paper[13] is based on Shannon's entropy. If the assumptions of the previous section are considered and s_{c_j} denote a set of elements of class j on the universal set X in an interval of R_j (it is a subset of the universal set X), the fuzzy entropy of a fuzzy rule (R_j) will be calculated by the following steps.

1) Calculating the match degree Dj of the elements of class j with the fuzzy set $\widetilde{A}$ for each rule by (7).

$$D_j = \sum_{x_p \in S_{c_j}} \mu_{\widetilde{A}}(x_p) \Big/ \sum_{x_p \in X} \mu_{\widetilde{A}}(x_p), (7)$$

2) Computing the fuzzy entropy $FE_{C_j}(\widetilde{A})$ for the elements of class j in an interval of $R_j$by (8).

$$FE_{C_j}(\widetilde{A}) = -D_j \log_2 D_j, (8)$$

3) Specifying the fuzzy entropy $FE(\widetilde{A})$ for the elements within the R_j's interval by (9).

$$FE(\widetilde{A}) = \sum_{j=1}^{j=M} FE_{C_j}(\widetilde{A}), (9)$$

III. PROPOSED HIERARCHICAL CLASSIFIER

As discussed above, one of the problems of Mamdani FRBCS is the lack of accuracy which is due to the inflexibility of the concept of linguistic variables[3]. One of the most important reasons of this inflexibility can be attributed to the rigid partitioning of the input-output spaces[14]. In order to improve the accuracy of FRBCS a new hierarchical classifier, which is named Halving Hierarchical Fuzzy Rule Based Classifier (HH-FRBC) is proposed in this paper. The aim of the proposed classifier is to use a more flexible KB. To serve this purpose a new KB which is named Halving Hierarchical Knowledge base (HHKB) is presented. HHKB is composed of Halving Hierarchical Data Base (HHDB) and Halving Hierarchical Rule Base (HHRB). The details of the new hierarchical classifier will be illustrated in the next subsections.

A. HHDB

The linguistic terms of the HHDB are formed in different levels hierarchically. In each level the DB will be updated using the information of the linguistic terms of the previous level. The number of linguistic terms in the fuzzy partitions of the first level is 2. Fig. 2 shows the membership functions (MFs) which are used in the first level of HHDB.

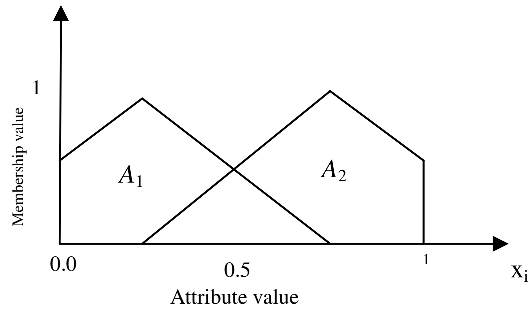

Fig. 2. the membership functions which are used in the first level of HHDB. A_1, A_2

In the proposed HHDB, the MFs of each new level are formed by halving the interval of the MF in previous level . Fig.3 illustrates the process of creating the DB of the second level by means of the DB of the first and Fig.4 illustrates the process of creating the DB of the third level by means of the DB of the second level.

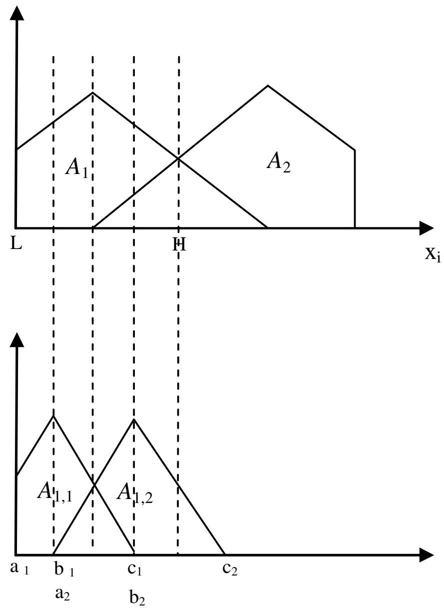

Fig. 3. mapping between terms from first level to second level

If the interval values of the MF of previous level are considered as L and H, the parameters of the MFs of new level of HHDB are calculated as follows: If one of the interval values of the MF which is considered to be halved equals to one of the interval values of the variable (Fig.3), the parameters of the MFs of new level are calculated as Eq. 10 or Eq. 11. Eq. 10 and Eq.11 are used for halving A_1 and A_2 respectively.

$$\begin{cases} a_1 = L, b_1 = L + d, c_1 = H - d \\ a_2 = L + d, b_2 = H - d, c_2 = H + d \end{cases}, (10) \quad \text{or}$$

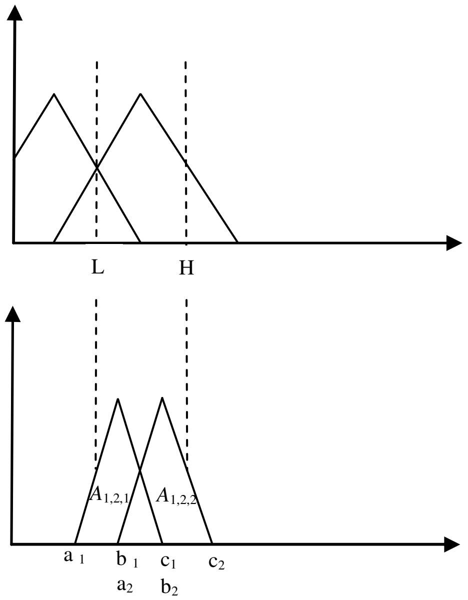

Fig. 4. mapping between terms from second level to third level

$$\begin{cases} a_1 = L - d, b_1 = L + d, c_1 = H - d \\ a_2 = L + d, b_2 = H - d, c_2 = H \end{cases},(11)$$

if the interval values of the MF which is considered to be halved does not equal to the interval values of the variable (Fig. 4)Eq.12 is employed.

$$\begin{cases} a_1 = L - d, b_1 = L + d, c_1 = H - d \\ a_2 = L + d, b_2 = H - d, c_2 = H + d \end{cases},(12)$$

Where, $d = (H - L)/2$. Similarly, the MFs of each new level can be constructed by the MFs of previous level and HHDB will be formed.

B. HHRB

The main objective of developing an HHRB is to model the problem space in a more accurate way based on the data distribution. The HHRB is formed by the MFs created in different level of HHDB. All possible fuzzy rules of the new level can be created as follows. For each existing rules of k_{th} level, two fuzzy rules of $(k+1)_{th}$ level will be created by halving the MFs related to one of the variable of kth level rule , and forming the rules corresponding to new MFs. It's worth mentioning that, for each rule of kth level, the best variable will be chosen for halving. Fig.5 and Fig.6 show all possible fuzzy rules of first and second level respectively.

C. The creation of linguistic values for rules of different levels

In the proposed system, like the other FRBCs, the antecedent of fuzzy rules contains linguistic variables and values, but in the proposed system, the linguistic values are obtained based on the rule's level. For example, consider R in Fig. 7(a), which is a rule in the third level. This rule halving is occurred along the x_1 for all three levels. The linguistic value of x_1 will be determined as follows:

1) In the first level, after halving initial space along x_1 (i.e. $A_1$and A_2), the second linguistic value which is equivalent to $A_2$is selected. Fig. 7(b).
2) In the second level, the subspace $A_2$of Fig. 7(b) which is shown by bold line is halved (i.e. A_{21} and A_{22}) and the first linguistic value which is equivalent to $A_{2,1}$ is selected. Fig. 7(c).
3) In the third level, the subspace A_{21} of Fig. 7(c) which is shown by bold line is halved (i.e. A_{211} and A_{212}) and the second linguistic value which is equivalent to $A_{2,1,2}$ is chosen. Fig. 7(d).

Consequently, linguistic value x_1 in R1 is represented by $A_{2,1,2}$. Bearing in mind that in this rule, x_2 is not selected to halve, don't care is chosen for this variable and x_2 is omitted from the rule representation. Therefore R will be represented as follows:

If x_1= $A_{2,1,2}$ then class=*C1*.

D. Proposed algorithm

HHFRBC creates rules in a hierarchical way by means of HHDB and HHRB. Fig.8 illustrates the outline of the proposed algorithm. In the following section, different steps of the proposed method are clarified through an example.

It should be mentioned that, the rule whose fuzzy entropy should be calculated at the beginning of the algorithm is the initial space of the problem.

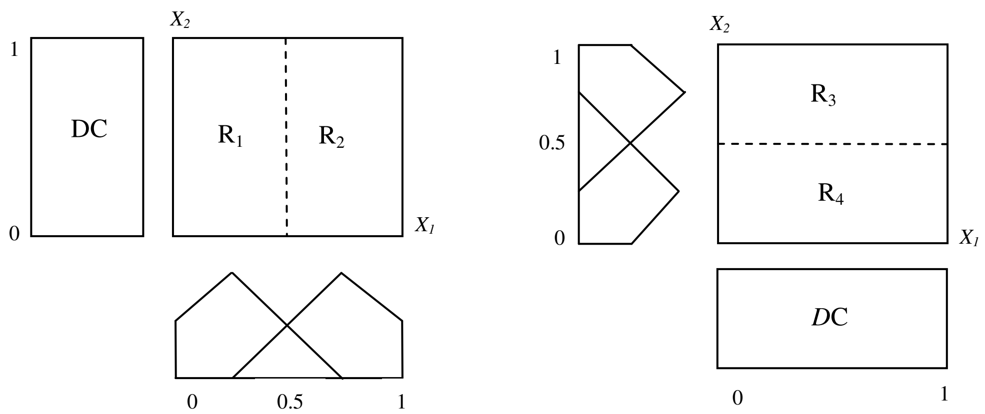

Fig. 5. All possible first-level rules: R_1 and R_2 are created by halving the initial space along the X_1, R_3 and R_4 are created by halving the initial space along the X_2

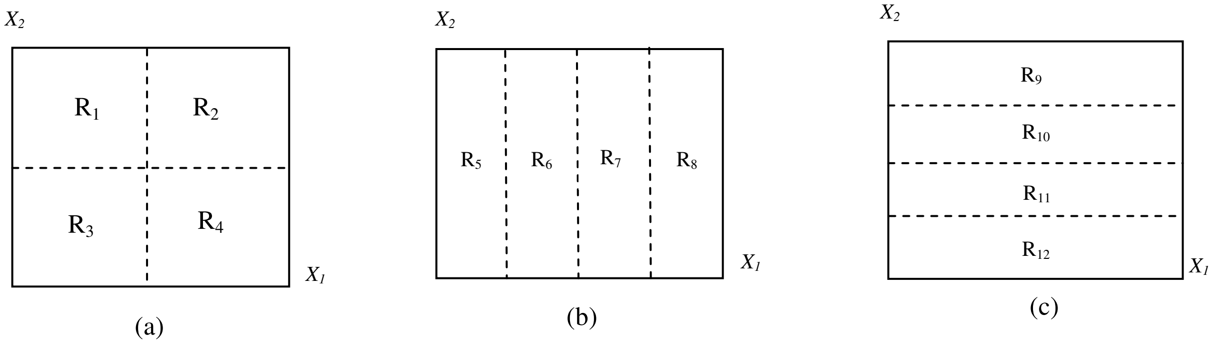

Fig.6. All the possible unique second-level rules: (a) Second-level rules which are created by halving R_1 and R_2 of the Fig. 2 along the X_2or R_3 and R_4 of the Fig. 2 along the X_1; (b) Second-level rules which are created by halving R_1 and R_2 of the Fig. 2 along the X_1; (c) Second-level rules which are created by halving R_3 and R_4 of the Fig. 2 along the X_2.

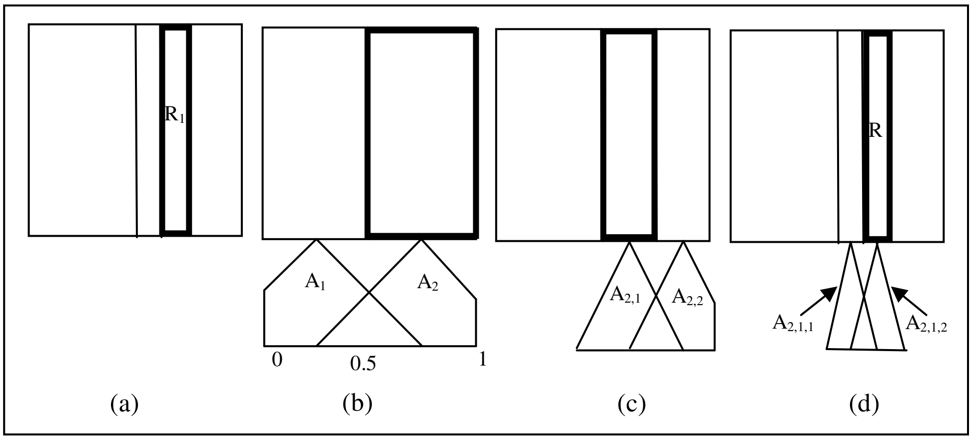

Fig. 7. (a) $R_1$is a rule in the third level *and* x_1 *is* chosen in all its three levels to be halved (b), (c) and (d) are the steps of creating the linguistic value corresponding to $x_1$of rule R_1

```
While (1)
{
  For each existing fuzzy rule
    {Calculate the Fuzzy Entropy by eq.9}
   If there are not any rules whose entropy is more than a threshold
    {Exit}
   Else
 Step0:  Select the rules whose entropy is more than a threshold
   For each selected rule
     {
Step1:  Compute the Fuzzy Entropy by halving each variable
Step2:  Select the variable that causes the best entropy
Step3: Exchange the selected rule for two created rules produced by halving it along the variable which causes the best entropy
     }
}
Prune the final rule set by validation set
```

Fig. 8. The outline of the proposed algorithm

Suppose Fig.9 shows the initial space of the problem with some data of two classes and we want to show different steps of the proposed method to classify it. For this example the threshold is considered as 0.

Step0: since the FE of the initial space is more than zero, this rule (i.e. the initial space) is selected for this step.

Step1: At this step, the FE produced by halving each variable is calculated by Eq. 13.

$$FE = (n_1/N)*FE(R_2)+(n_2/N)*FE(R_3), \quad (13)$$

Where, N is the number of total dataset in R_1, n_1 and n_2 are the number of samples belongs to the R_2 and R_3 respectively. Fig.10 and Fig.11 illustrate the process of halving R_1 along the first and second variable.

Step2: Since the FE produced by halving the first variable (Fig.10) is less than the fuzzy entropy produced by halving the second variable (Fig.11), first variable is chosen as the best one.

Step3: at this step, R_1 will be exchanged for R_2 and R_3which can be seen in Fig.10.

At this point FE will be calculated for R_2 and R_3 by Eq. 9 and step0 to step3 is examined for them. This process is iterated until the FE of all rules becomes zero. After two times, the final rule set which is shown in Fig. 12 will be created.

At last, due to the stopping criterion (i.e. threshold=0 for all rules) which is used in this paper, the final rule set extracted by the method may over fit the data. To overcome this over fitting problem, the final rule set will be pruned by means of validation set (i.e. with remaining a subset (1/3) of training data to use for pruning). In such a manner that, after creating the final classifier, each rule can be removed if the accuracy of the classifier does not decrease on validation set.

IV. EXPERIMENTAL STUDY

In our experiments, Five- fold cross validation approach (i.e. five partitions are used for training and test sets, 80% for training and 20% for test) is applied in all of the experiments. After that, for each dataset, the average result of the five partitions is considered. To detect significant differences among the results produced by the studied method, some statistical tests need to be performed. There are different tests for multiple comparison over multiple datasets[15, 16]. Since there is not any assumption for normality in the nature of our problem, Friedman and Iman and Dovenport tests [17, 18] are chosen to compare multiple classifiers. Holm method [19] is employed as a post processing method to compare the best ranking classifier with the remaining ones. .(A complete description of these tests and the software used for its calculation can be found in the http://sci2s.urg.es/sicidmwebsite). In this section, first, the data sets which are used in this paper will be presented before examining the effect of pruning in the HH-FRBC. In the following stage, a comparison between the results of HHFRBC and some other classifiers will be conducted.

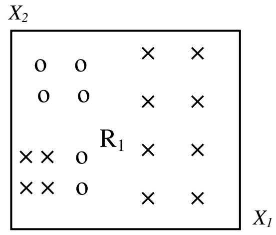

Fig. 9. initial space of the problem which is considered to classify

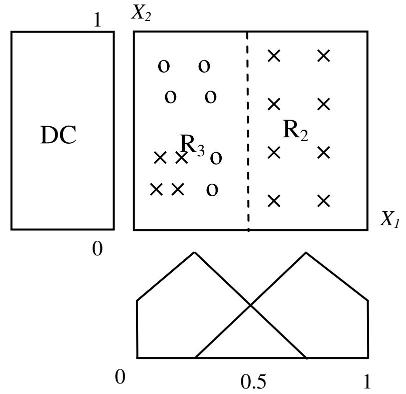

Fig. 10. Process of halving initial space along the first variable

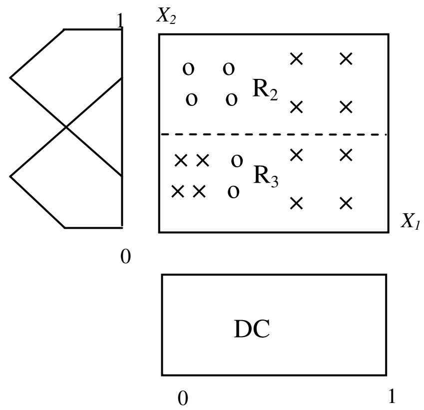

Fig. 11. process of halving initial space along the second variable

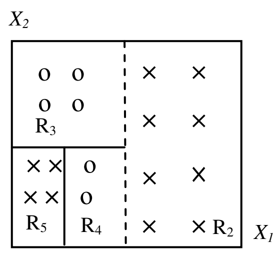

Fig. 12. the final rule set

The effect of hierarchical structure is analyzed by comparing its performance with the one obtained by the simple Chi algorithm in standard classification[9]. Furthermore, to examine the behaviour of the proposed hierarchical fuzzy rule-based classifier, a comparison between the results of HFRBCS [7] and the proposed structure will be carried out.

A. Experimental setup and the effect of pruning

In this paper, 8 data sets from UCI machine learning repository[20] have been selected. These data sets are summarized in Table. 1 along with their name, number of examples (#Ex), attribute (#Atts) and classes (#Class). Besides, the results of LDWPSO algorithm were obtained by running Keel.2 software [21]. Most of the approaches to constructing final rule sets involve greedy heuristics (such as fuzzy entropy reduction) that may over fit the training data and cause to low accuracy in test data. To overcome the problem of over fitting, some pre-pruning and post-pruning methods are employed. . As earlier mentioned, a post-pruning method is used in this paper and the effect of this method on accuracy and the number of obtained rule is examined in this section. Table. 2 shows that test accuracy is increased slightly after pruning the final rule set and number of rules is decreased extremely.

TABLE I. Data description

Dataset	Number of attributes	Number of classes	Number of examples
Breast cancer	9	2	699
Wine	13	3	178
Iris	4	3	150
Vehicle	18	4	846
Bupa	6	2	345
yeast	8	10	1,484
Page-blocks	10	5	548
Ecoli	7	8	336

B. Total analysis of HH-FRBC

In order to show the advantages of HH-FRBC, the accuracy of HH-FRBC is compared with two FRBCs namely Chi3 and HFRBCS. Table. 3 and Table. 4 compare the results of HH-FRBC and other classifiers on training and test data set respectively. Table. 5, in turn, shows the number of rules obtained with each method .The average results in Table. 6 prove that HH-FRBC outperforms other algorithms on training data sets. The results of Table. 6 illustrate that the highest value of test accuracy is associated with HH-FRBC. The average number of rules of the proposed method is less than Chi3 and HFRBCS which is extracted from Table. 5.

TABLE III. Training results of HHFRBC in comparison with other classifiers

Dataset	HH-FRBC	Chi3	HFRBCS
Breast cancer	0.9806	0.9806	**1**
Iris	**0.9933**	0.9373	0.9493
Vehicle	0.9101	0.6587	**0.9366**
yeast	**0.9871**	0.2963	0.6399
Bupa	**0.9884**	0.60	0.8424
Page-blocks	**0.9661**	0.9260	0.9514
Wine	0.9943	0.9874	**0.9994**
Ecoli	**0.9940**	0.7833	0.9326
average	**0.9767**	0.7712	0.9064

TABLE II. The accuracy and the number of rules of the proposed method with and without employing pruning method

Dataset	accuracy without pruning	accuracy with pruning	The number of rules without pruning	The number of rules with pruning
Breast cancer	**0.9617**	**0.9661**	45	7.8
Iris	0.9466	**0.9533**	36.2	7.8
Vehicle	0.6733	**0.7124**	1895	104.6
yeast	0.6486	**0.6668**	2412	262.4
Bupa	0.7130	**0.6927**	259.6	48
Page-0blocks	0.9451	**0.9425**	2802	96.8
Wine	0.8742	0.8914	25.4	8.2
Ecoli	0.8358	0.8089	180	31.8
average	0.8247	**0.8292**	956.9	70.92

TABLE IV. Test results of HHFRBC in comparison with other classifiers

Dataset	HH-FRBC	Chi3	HFRBCS
Breast cancer	**0.9661**	0.9121	0.9107
Iris	**0.9533**	0.94	0.9267
Vehicle	**0.7124**	0.6077	0.6726
Yeast	**0.6668**	0.2918	0.5748
Bupa	**0.6927**	0.5787	0.6259
Page-blocks	**0.9425**	0.9161	0.9270
Wine	0.8914	**0.9268**	0.9160
Ecoli	0.8089	0.7833	**0.8719**
average	**0.8292**	0.7445	0.8032

TABLE V. Average number of rules for HHFRBC in comparison with other classifiers

dataset	HH-FRBC	Chi_3	HFRBCS
Breast cancer	7.8	224	216.3
Iris	7.8	16.6	18.4
Vehicle	104.6	227.8	2041
yeast	262.4	25.4	216
Bupa	48	43.4	120
Wine	8.2	120.2	144.9
Page-blocks	96.8	19.6	39.4
Ecoli	31.8	43.5	158.9
Average	70.92	90.06	369.36

In addition, Friedman and Iman-Dovenport tests are used to see if there are significant differences in the results. The results of applying these tests are shown in Table. 6.

Because the P-values produced by Friedman and Iman-Davenport tests in Table. 7 are less than 0.05, it can be concluded that there are significant differences among the results obtained by multiple classifiers. In other words, obtained p-values inform us of the rejection of the null hypothesis (i.e. equality of means). Furthermore, the ranking which is computed using a Friedman test of the 5 algorithms is illustrated in Fig.13. The value given to each method is acquired by assigning a position to each algorithm depending on its performance for each data set. The algorithm with the best accuracy on a specific data set receives the first ranking (value 1); then, the algorithm with the second best accuracy receives rank 2, and so forth. This assignment is done for all data sets. At last an average ranking is calculated as the mean value of all ranks.

Fig.12 illustrates that the best ranking method is HH-FRBC. Now the best ranking method (HH-FRBC) is selected as control method and a post-hoc test (i.e. Holm test in this case) is employed to find the algorithms that reject the equality hypothesis. Table.7 shows the result of this test. The corresponding p-values associated with each comparison are obtained by Normal distribution. To show whether the hypothesis is rejected in favor of the control method or not, obtained P-values can be compared with the associated α/i in the same row of the table. If the p-value is less than associated α/i, it shows that control method rejects corresponding hypothesis. Building on the above discussion, the results of Table. 7 show that, the proposed method rejects Chi3 approaches.

Although the proposed method achieves a higher ranking than HFRBCS, this is not enough to reject this hypothesis. Therefore, it may be concluded that both approaches have a similar performance. Obtained results in Table. 7 demonstrate goodness of the proposed hierarchical structure, because HH-FRBC outperforms simple Chi algorithm in standard classification. Although HH-FRBS cannot reject HFRBCS, there is tremendous difference between the number of rules obtained by these two methods which result from the proposed hierarchical structure.

V. CONCLUDING REMARKS

In this paper a new hierarchical fuzzy rule-based classifier which is called HH-FRBC is proposed which is based on binary tree decomposition.

TABLE VI. Results of the Friedman and Iman-Davenport tests for comparing performance of the HH-FRBC in all data sets

method	P-value
Friedman	0.0075
Iman-dovenport	0.0023

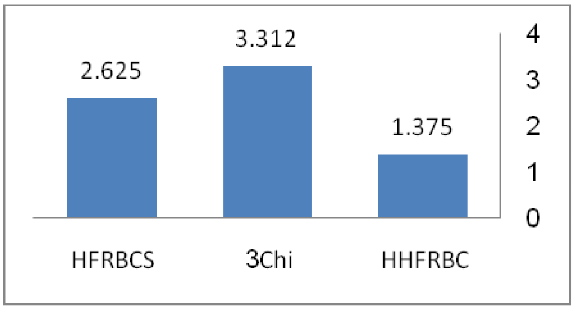

Figure 13.Ranking of different classifiers obtained by Friedman test

Table VII. Results of Holm test. PHH-FRBC is the control method

i	Algorithm	P	α/i	Hypothesis(α=0.05)
2	HFRBCS	0.11384629800665803	0.025	Not Rejected
1	Chi3	0.014255291355700033	0.05	Rejected for PHH-FRBC

each partition and a pruning method is employed in final rule set to overcome the problem of over fitting. The main objective of considering hierarchical structure for rules is generating fine and coarse fuzzy subspaces simultaneously to cover the problem space properly. In the experimental study, our method is compared with a simple Chi algorithm in standard classification and the last version of hierarchical fuzzy rule base classification system (i.e. HFRBC) and it is shown statistically that the performance of proposed classifier is better than most of them on most of the datasets on one hand and on the other hand, the number of rules obtained by the proposed method is less than other fuzzy rule based classifier systems, which can proves the interpretability of the proposed classifier. The crux of the argument is that the proposed method provides trade-off between the interpretability and accuracy.

[1] A. Bardossy and L. Duckstein, Fuzzy Rule-Based Modeling with Application to Geophysical, Biological and Engineering Systems CRC press, 1995.

[2] L. A. Zadeh, "Outline of a new approach to the analysis of complex systems and decision Outline of a new approach to the analysis of complex systems and decision processes," IEEE Transactions on Systems, Man and Cybernetics SMC-3, pp. 28-44, 1973.

[3] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning," Information Sciences, vol. 8, pp. 199-249, 1975.

[4] O. Cordon, F. Herrera, F. Hoffmann, and L. Magdalena, Genetic Fuzzy Systems Evolutionary Tuning And Learning Of Fuzzy Knowledge Bases: World Scientific Publishing Co. Pte. Ltd, 2001.

[5] A. Bastian, "How to handle the flexibility of linguistic variables with applications," International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems vol. 2, pp. 463-484, 1994.

[6] O. Cordón, F. Herrera, and I. Zwir, "Linguistic Modeling by Hierarchical Systems of Linguistic Rules," IEEE Transaction on fuzzy systems, vol. 10, pp. 2-20, 2002.

[7] A. Fernández, M. J. d. Jesus, and F. Herrera, "Hierarchical fuzzy rule based classification systems with genetic rule selection for imbalanced data-sets," International Journal of Approximate Reasoning, vol. 50, pp. 561–577, 2009.

[8] A. Fern´andez, M. J. d. Jesus, and F. Herrera, "Analysing the Hierarchical Fuzzy Rule Based Classification Systems with Genetic Rule Selection," 4th International Workshop on Genetic and Evolutionary Fuzzy Systems (GEFS), 2010.

[9] Z. Chi, H. Yan, and T. Pham, "Fuzzy algorithms with applications to image processing and pattern recognition," World Scientific, 1996.

[10] O. Cord´on, M. ı. J. e. d. Jesus, and F. Herrera, "Genetic Learning of Fuzzy Rule-Based Classification Systems Cooperating with Fuzzy Reasoning Methods," International Journal of Intelligent Systems, pp. 1025-1053, 1998.

[11] H. Ishibuchi and T. Nakashima, "Effect of rule weights in fuzzy rulebased classification systems," IEEE Transactions on Fuzzy Systems, vol. 9, pp. 506-515, 2001.

[12] H. Ishibuchi and T. Nakashima, "Techniques and applications of genetic algorithm-based methods for designing compact fuzzy classification systems," in Fuzzy theory systems: techniques and applications. vol. 3: Gulf Professional, 1999, pp. 1081-1109.

[13] Hahn-Ming Lee, Chih-Ming Chen, Jyh-Ming Chen, and Y.-L. Jou, "An Efficient Fuzzy Classifier with Feature Selection Based on Fuzzy Entropy," IEEE Transaction on fuzzy systems, man , and cybernetics-part B: cybernetics, vol. 31, 2001.

[14] A. Bastian, "How to handle the 7exibility oflinguistic variables with applications," International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 2, pp. 463–484, 1994.

[15] J. Demˇsar, "Statistical Comparisons of Classifiers over Multiple Data Sets " Journal of Machine Learning Research, vol. 7, pp. 1-30, 2006.

[16] S. Garc´ıa and F. Herrera, "An Extension on “Statistical Comparisons of Classifiers over Multiple Data Sets” for all Pairwise omparisons," Journal of Machine Learning Research vol. 9, pp. 2677-2694, 2008.

[17] M. Friedman, " A comparison of alternative tests of significance for the problem of m rankings," Annals of Mathematical Statistics, vol. 11, pp. 86-92, 1940.

[18] R. L. Iman and C. i. S. J. M. Davenport, "Approximations of the critical region of the friedman statistic , Part A – Theory Methods," vol. 9 pp. 571–595, 1980.

[19] S. Holm, "A simple sequentially rejective multiple test procedure," Scandinavian Journal of Statistics, vol. 6, pp. 65–70, 1979.

[20] C. Blake, E. Keogh, and C. J. Merz, "UCI repository of machine learning database," 1998.

[21] J. Alcalá-Fdez, A. Fernandez, J. Luengo, J. Derrac, S. García, L. Sánchez, and F. Herrera, "KEEL Data-Mining Software Tool: Data Set Repository, Integration of Algorithms and Experimental Analysis Framework," Journal of Multiple-Valued Logic and Soft Computing, 2010.